CHAPTER 9

CONCLUSION AND FUTURE WORK

9.1 CONCLUDING REMARKS

This thesis investigated five new approaches for text-independent speaker identification by exploiting the use of cepstral features. Dynamic cepstral coefficient feature is obtained by incorporating pitch frequency information in Mel scaled cepstral data. Since, speaker identification process is the replica of human hearing mechanism, feature deals with human auditory system is clubbed with pitch based vocal tract feature and produced good speaker identification accuracy. The block truncation and wavelet based sub band processing techniques helps to eliminate the major limitation of GMM, i.e., the requirement of large amount of speech data to train the mixture model. Most of the researches in speaker identification field concentrated on monolingual speech data for speaker identification, whereas this thesis investigated multilingual speaker identification with speech signal produced in English and some of the Indian languages such as Tamil, Hindi and Telugu. A quadrilateral structured new filter bank is introduced to show its superiority over triangular and Gaussian shaped Mel filter banks in speaker identification performance.

Block truncated DCT effectively decorrelates filter bank log energies better than conventional DCT. When the time-frequency localized wavelet sub band processing is followed by block truncated DCT technique, the speech feature is very much suitable for speaker identification in noisy
condition specifically for narrowband noise. Multi level wavelet decomposition splits the energy spectrum into bands, thus the influence of the narrowband noise affected bands is restricted in rest of the bands.

The work in the thesis used fusion of complementary information for improved speaker identification. Standard complementary information fusion approaches, fuse the speech signal low frequency features with its complementary high frequency features such as, fusion of Mel MFCC feature with IMFCC feature. But in this MFCC and IMFCC fusion technique, the improvement in identification accuracy is only a little against identification accuracy with only MFCC feature. The work presented in this thesis used fusion of vocal tract feature with its complementary aural feature and found to exhibit improvement in speaker identification accuracy.

Unlike popular monolingual speech database, standard multilingual speech databases are not readily available for use in speaker identification system. Hence, to carry out multilingual speaker identification experiments, an indigenous multilingual speech database - MEPCO speech database is created in the college acoustic environment with students who can speak four different languages. Multilingual speaker identification on MEPCO speech database made of native speakers of Tamil language was conducted and the speaker specific information acquiring ability of MFCC against few other cepstral coefficients are examined. The performance of speaker identification is evaluated when same language is used for testing also when different languages are used for testing. Moreover, the identification performance is tested with English, Tamil, Hindi and Telugu language utterances generated synthetically. A comparison is made for multilingual speaker identification system between GMM and ELM model methods.

In designing Mel scaled filter bank for extracting MFCC feature, conventionally triangular filters are equally spaced along the Mel frequency
scale to yield the triangular Mel scale filter bank structure. This filter bank structure is widely used in most of the speaker recognition researches. But the problem associated with this bank structure is that, adjacent filters in the filter bank are having sharp transitions which may cause some of the useful information left out in the boundary region. To overcome this issue, Gaussian shaped Mel scale filter bank structure was introduced by the research community. But the problem with this Gaussian filter structure is over encompassing of adjacent filters. To address this issue, a filter structure which neither under fits nor over fits is required. The Quadrilateral filter structure proposed in this thesis meets this requirement and works for effectively yielding the MFCC. Some of the speaker dependent information, which are scattered around the lower frequency range of each filter bin is not given any priority during design of Triangular and Gaussian filter bank structure. Hence, a new filter bank which covers more low frequency range for every filter bin is designed and the resultant filter bank structure is named as Quadrilateral Filter bank. For designing any filter bank its center frequency, ERB is essential to determine the lower, upper and intermediate frequencies of every filter bin.

In this thesis, text independent speaker identification system under noisy environment is presented using RASTA-MFCC as feature vector and GMM-UBM is chosen as the modeling since it is highly resistive to imposter attack. A new Quadrilateral filter bank structure is designed and its performance is found to be better than conventional filter banks. The added advantage of GMM-UBM is that it requires minimum number of mixture components for achieving good identification results. Experimental results shows that the RASTA-MFCC features which uses Quadrilateral filter banks are more robust to noisy environment than triangular and Gaussian filter banks. The UBM adaptation is faster than GMM training. The quality of UBM is much better than GMM when small training segments on the order of 2 to 5 seconds is used.
The complexity of the proposed approaches are explained as follows: While many existing systems for speaker identification achieved good performance in relatively constrained environments, performance invariably deteriorates in constrained environments like i) noisy environment ii) limited data condition iii) multilingual condition iv) mimicry by human and v) multi-speaker condition. Every unconstrained environment poses biggest challenge in speaker identification accuracy. Noisy environment warrants the accurate estimation and removal of noise. Limited data condition adds the complexity of data scarcity in modeling speaker parameters. Enrolment in one language and testing in another language is challenging in Multilingual condition. Imitating like another one enrolled person in speaker identification through mimicry increases system complexity.Moreover identifying a person or separating the speakers in the multi-speaker environment is the complex task speaker identification. All the complexities are added and tested in the proposed speaker identification approaches.

9.2 SCOPE FOR FURTHER RESEARCH

In the speaker identification system presented in this thesis, wavelet based sub band processing followed by block truncation was introduced. Pitch frequency based speaker specific features are extracted. Because of this step it was possible to address the limited data speaker identification under noisy condition and to increase the identification accuracy. This work can still be improved in the future by replacing wavelet sub band with wavelet packet sub bands, which will give importance to both low frequency and high frequency components of speech signal. Moreover, model score fusion could be done using regression based fusion techniques.

Secondly, for multilingual speaker identification system, along with the features used in chapter 4 and 5, other features such as language specific intonation feature and semantic features could be used to improve the speaker
identification efficiency. Beside that, work can be contributed to advance the ELM algorithm in order to improve the speaker identification accuracy along with its existing fast computational speed. Apart from this, a widely accepted multilingual speech database could be generated for use in multilingual speaker identification by researchers around the world, considering the application of secured data accesses internationally by the people who use to talk in different languages. This thesis has introduced the advantageous quadrilateral filter bank structure, which performed better than widely used triangular shaped Mel scale filter bank and little used Gaussian shaped Mel scale filter bank. Still to increase the identification accuracy using speaker specific data, new structures of Mel scale filter bank could be explored. An integration of low level and high level wavelet packet sub band processed, block truncated features could also improve the speaker identification system performance.

Considering the scope for future research in the context of speaker identification task as a helping hand for mankind, the proposed wavelet processing and block truncation based narrowband noise robust speaker identification approach and aural and vocal tract feature based speaker identification approach can be expanded with additional features which have the capability to capture most of the speech and speaker attributes to help assess the pathological voice quality for assessing human voice who are suffered by dysphonia and dysarthric impairments and are getting medication for voice quality improvement. Pathological voice assessment using speaker identification system aims to provide a method, suitable for keeping track of the evolution of the patient’s pathology. Since this system is easy-to-use, fast, non-invasive for the patient, and affordable for the clinicians, this approach of voice assessment using speaker identification system will be a real helping hand for mankind.